BIS for Music Streaming Platform

Final Presentation

Group 7
Amanda Raj Shrestha   122245
Min Khant Soe   122277
Sunny Kumar Tuladhar   122336
Win Win Phyo   122314

Objective

To create **Business Intelligence System** for **music streaming platform.**

Our main goal is to create an Business intelligence system to increase the **customer engagement** to the **platform** i.e *number of listens*

This in turn will increase the revenue of the platform and give insights to its users and its artists that help increase the listens.
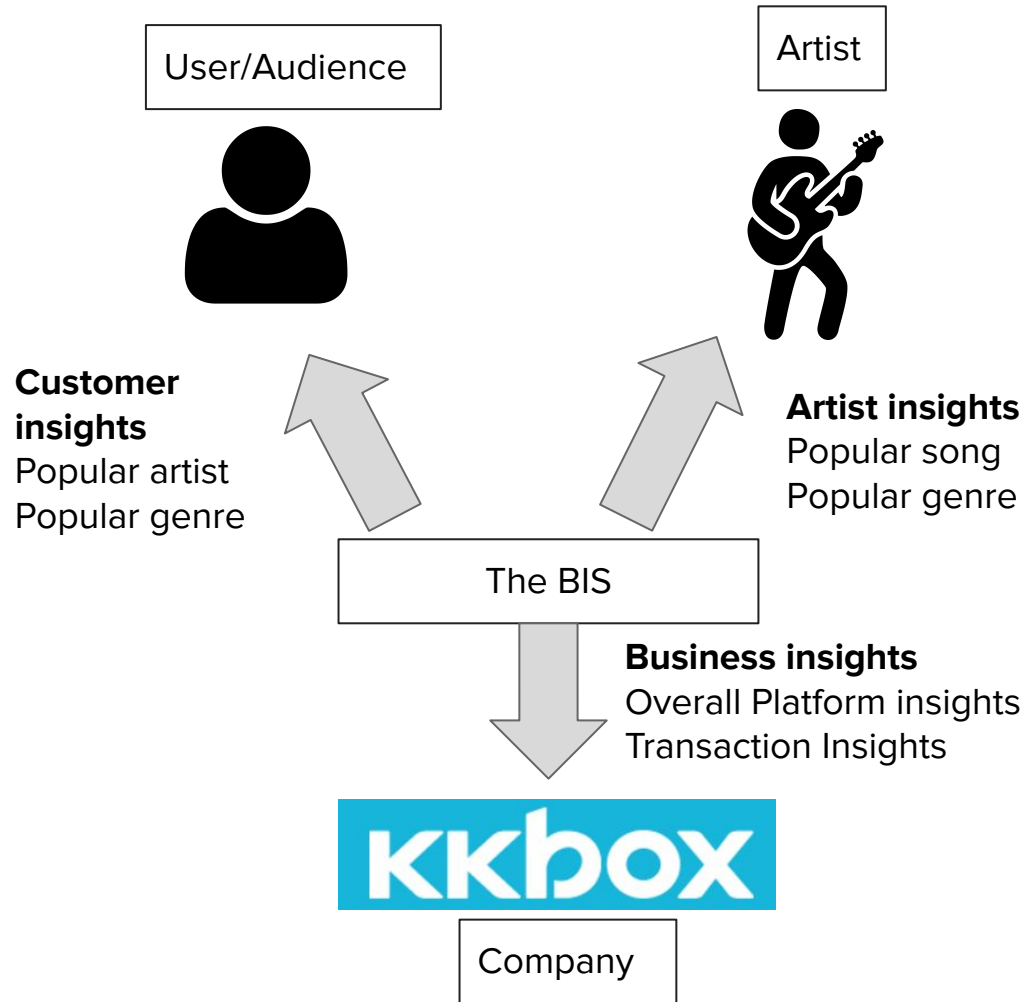
End Users

User/Audience: The users who listen to music and purchase subscriptions. Main source of income.

Artists: The musicians who upload their music which the users listen to

Company: The company who owns the music streaming platform

User/Audience

Artist

Customer insights
Popular artist
Popular genre

Artist insights
Popular song
Popular genre

The BIS

Business insights
Overall Platform insights
Transaction Insights

KKbox

Company

Tasks

- **To visualize useful data and current trends from artists , songs and users**
- **To analyze data from visualization**
- **Create a prediction model to predict whether a user will like a song**

Dataset used

WSDM - KKBox's Music Recommendation Challenge

https://www.kaggle.com/c/kkbox-music-recommendation-challenge

WSDM - KKBox's Churn Prediction Challenge

https://www.kaggle.com/c/kkbox-churn-prediction-challenge/data?select=members_v3.csv.7z

KKBOX

🌐 kkbox.com

KKBox, stylized as KKBOX is a music streaming service developed in 2005 by KKBox Inc., a software company in Taipei, Taiwan. It is a part of Japanese Telecom Group, KDDI. Wikipedia

Launched : 2005

Users : 10 million+

Founder(s) : Chris Lin, et al .

Headquarters : Taipei , Taiwan

Founded : October 2004; 17 years ago

Owner : KKBOX

Software and Programming for implementation

- **Python** - for data preprocessing and Machine Learning models
- **Tableau** - for data visualization

Dataset Tables

- **User_Info** - Contains information about the users
- **Transactions_Info** - Contains information about the transaction/Subscription made by the users
- **Song_Info** - contains information about the song like length, genreid
- **Song_Extra_Info** - contains song names
- **User_Song_Relation_Info** - contains song and user relation, where the user listened to it and did they like it

Linked by User ID

Linked by Song ID

Data preprocessing

For Visualization,
- We checked the null values from all the data and filled them with relevant data.

- Since these data are not for machine learning tasks, the names of cities, genre category and language of songs are presented in number format. So, we have to manually fill text representation of them which were found by researching on the internet.
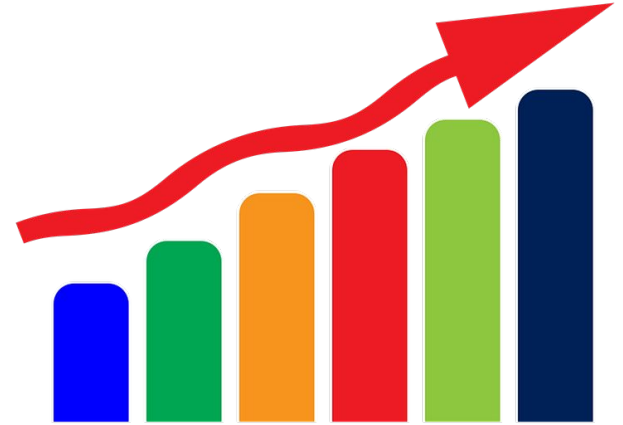
Data preprocessing

For Machine Learning
● All categorical data was converted to numerical using pandas library.

● The **user_info** and **song_info** was merged with **song_user_relations** using the **songID** and **userID**. Then 7 features were selected as shown below to be used for classification algorithm to run on.

	song_length	genre_ids	language	source_system_tab	source_screen_name	source_type	bd	city	gender
0	247640	342	3.0	3	8	3	53	1	2
1	197328	321	31.0	3	8	3	53	1	2
2	292571	342	3.0	3	8	3	53	1	2
3	226742	342	3.0	3	8	3	53	1	2
4	238132	335	3.0	3	8	3	53	1	2

Analytics Types

- **Descriptive**
    - User Age-group and Gender ratio (User Demographic)
    - Top performing songs/ genres/artists (Trends)
    - Timeline of Revenue generated and New subscriptions of users

- **Predictive**
    - Predict whether a user will like a song or not

- **Prescriptive**
    - Song Recommendation for the users.
    - Genre Recommendation for artists
    - Which artists to promote by the Streaming Platform

DASHBOARDS

Company

The Company dashboard aims to provide the company with the overall insight of the platform.

It aims to provide insight about

1. Customer demographic
2. Revenue visualisation
3. Artist/Genre/Song performance

Company

Customer demographic

- **Shows highest user demographic**
- **Ad companies which fit the highest demographic can apply**
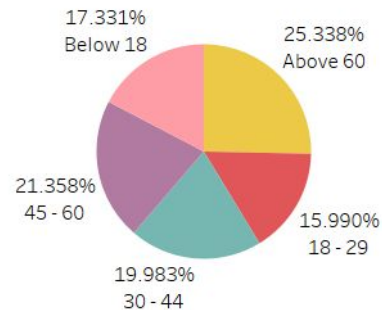
Total Users

6,769,473

Top 5 Cities

cities	
New Taipei	4,804,176
Kaohsiung	385,066
Taichung	321,016
Taipei	246,864
Taoyuan	210,429

Gender Percent per Age Group

Age Group	Gender female	male
Below 18	48.959%	51.041%
18 - 29	48.955%	51.045%
30 - 44	48.868%	51.132%
45 - 60	48.771%	51.229%
Above 60	49.026%	50.974%

Listeners' Age Group

17.331% Below 18

25.338% Above 60

15.990% 18 - 29

19.983% 30 - 44

21.358% 45 - 60

Gender
- female
- male

Age Group
- Above 60
- 18 - 29
- 30 - 44
- 45 - 60
- Below 18

User Population Count
183,311 — 301,829

Listeners' Gender Ratio

48.917% female
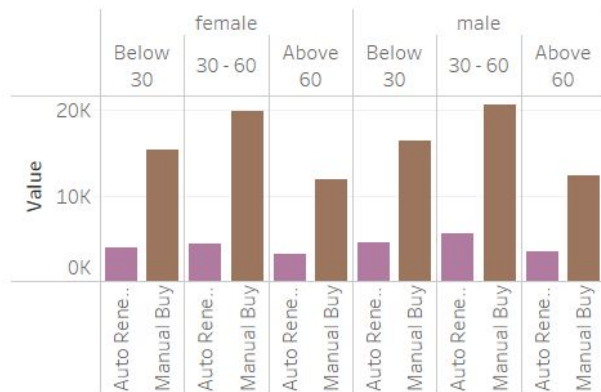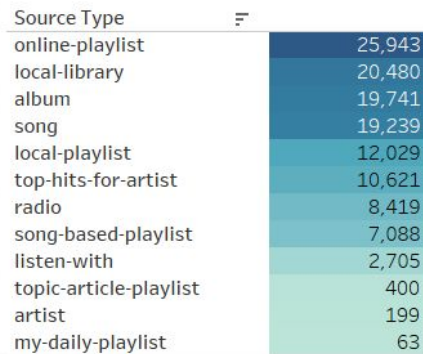
51.083% male
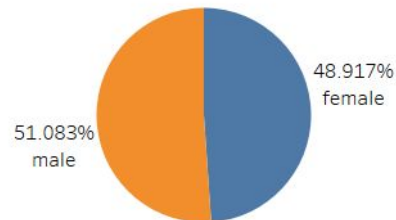
Company

Customer Behaviour

- **Plan which gender, age group to target the platform at**
- **Put trending songs in most popular source type**

Auto Renew & Manual Buy by Gender and Age Group

Source Type Used by Users to Search Songs

Source Type	
online-playlist	25,943
local-library	20,480
album	19,741
song	19,239
local-playlist	12,029
top-hits-for-artist	10,621
radio	8,419
song-based-playlist	7,088
listen-with	2,705
topic-article-playlist	400
artist	199
my-daily-playlist	63

Measure Names
- Auto Renew
- Manual Buy

Age Group
- Below 30
- 30 - 60
- Above 60

Plan Days
- 30
- 195
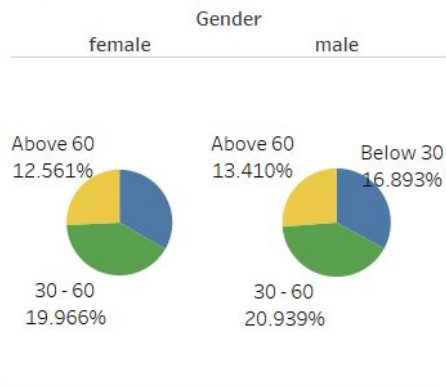- 410

Distinct count of us..
63   25,943

Subscription Cancellation by Gender and Age

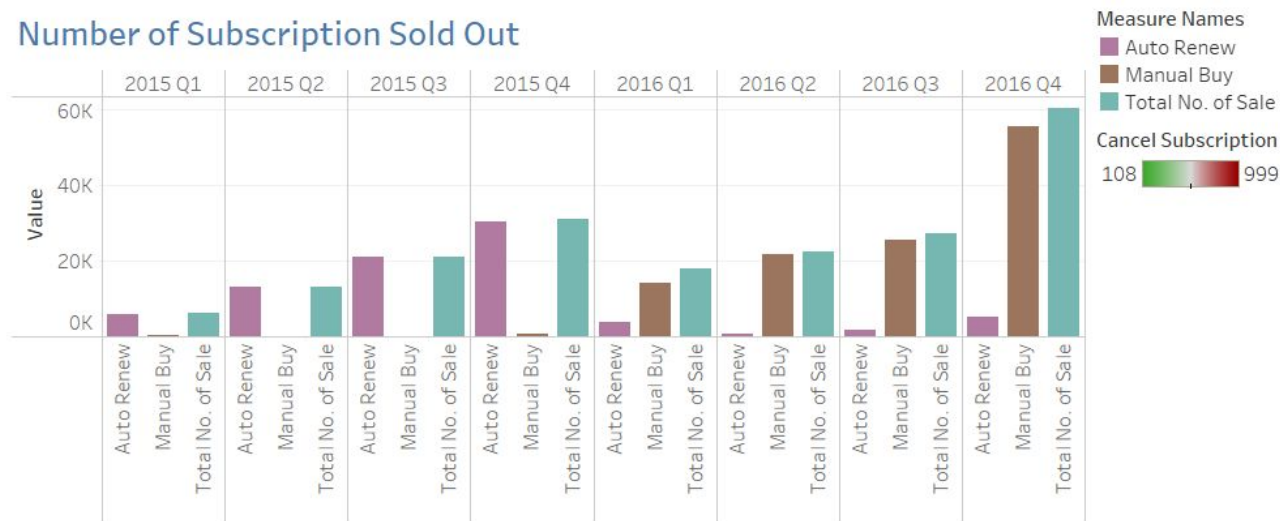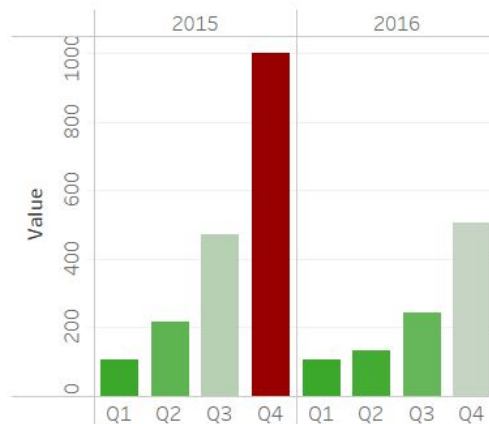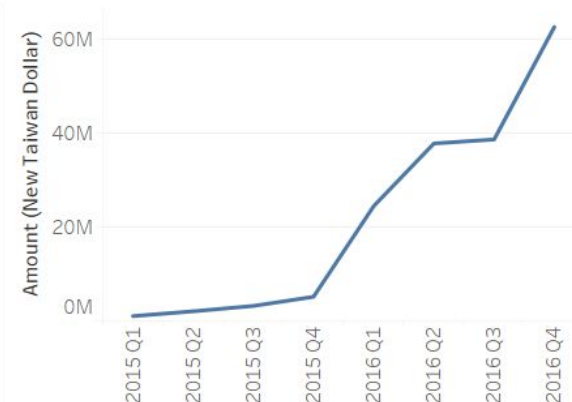## Top 3 Subscription Plans per Gender and Age Group

Company

Top songs/Artist and genre

- Analyse the top trending music on the platform.
- Recommend these trending music more for higher audience engagement on music platform.

Top 5 Songs

Name	Genre Ids	
帥到分手	Mandopop	10,612
告白氣球	Seasonal	295
	Mandopop	10,314
你, 好不好？(How Have You Been?)	Mandopop	9,742
小幸運 (A little happiness)	Seasonal	9,049
不該	Seasonal	42
	Mandopop	8,489

Top 5 Genres

Genre Ids	
Seasonal	1,779,253
Mandopop	660,203
Rock	170,908
R&B/Soul	152,342
444	115,986

Top 5 Artists

Artist Name	
Various Artists	149,727
五月天 (Mayday)	96,006
周杰倫 (Jay Chou)	107,386
林俊傑 (JJ Lin)	64,061
田馥甄 (Hebe)	57,772

Company

Top songs/Artist and genre

- **Analyse the top trending music on the platform.**
- **Recommend these trending music more for higher audience engagement on music platform.**
- **Put trending songs in most popular source type**

Top 5 Songs

Name	Genre Ids	
帥到分手	Mandopop	13,616
告白氣球	Mandopop	12,962
	Seasonal	566
你, 好不好? (How Have You Been?)	Mandopop	12,777
小幸運 (A little happiness)	Seasonal	12,574
不該	Mandopop	11,106
	Seasonal	81

Top 5 Genres

Genre Ids	
Seasonal	3,483,090
Mandopop	1,200,533
Rock	337,623
R&B/Soul	300,131
K-Pop	228,131

Source Type Used by Users to Search Songs

Number of Listeners: 378 — 2M

Source Type	
local-library	2,215,237
online-playlist	1,896,786
local-playlist	1,058,686
radio	477,558
album	471,944
top-hits-for-artist	259,123
song	218,048
listen-with	187,049
song-based-playlist	161,891
topic-article-playlist	9,645
artist	2,952
my-daily-playlist	378

Top 5 Artists

Artist Name	
周杰倫 (Jay Chou)	178,171
五月天 (Mayday)	172,710
林俊傑 (JJ Lin)	108,715
田馥甄 (Hebe)	99,814
陳奕迅 (Eason Chan)	71,858

User/Audience

User dashboard
- **Guide the listener to the top trends**

- **Show what the listener has been listening to lately**

Most Popular Songs - Overall

周杰倫 (Jay Chou) 告白氣球	周杰倫 (Jay Chou) 不該	Eric 周興哲 你, 好不好？ (How Have You Been?)	謝和弦 (R-chord) 謝謝妳愛我 (Thanks For Your Love)	陳奕迅 (Eason Chan)
周湯豪 (NICKTHEREAL) 帥到分手		林俊傑 (JJ Lin) 不為誰而作的歌 (Twilight)		Alan Walker Faded
田馥甄 (Hebe) 小幸運 (A little happiness)		五月天 (Mayday) 後來的我們 (Here| After| Us)		

User Listened

10,315   13,616

User Id

User 1

Listens

12   35
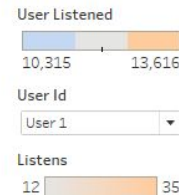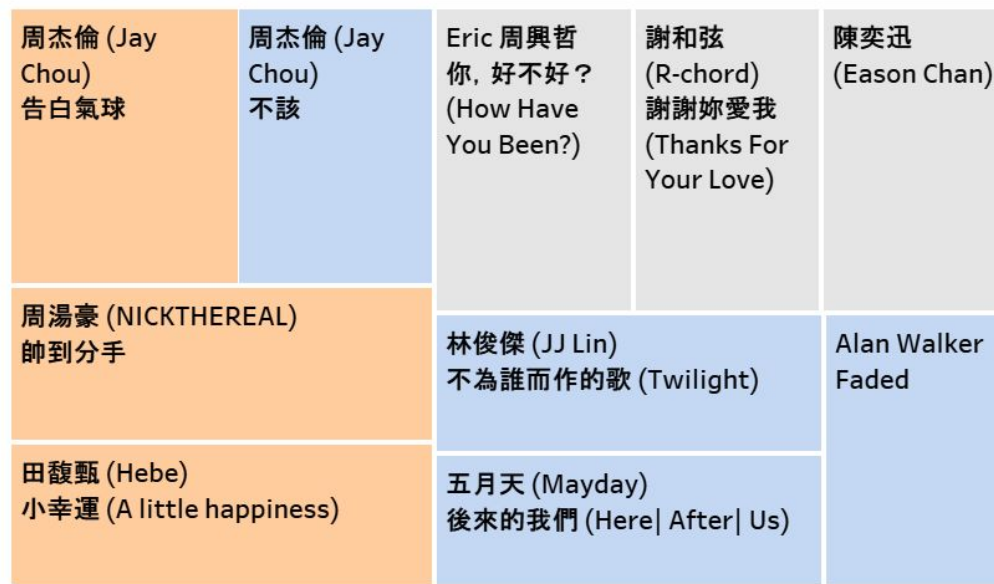
Most Popular Artists - Overall

Artist Name	
Various Artists	1
五月天 (Mayday)	2
周杰倫 (Jay Chou)	3
林俊傑 (JJ Lin)	4
田馥甄 (Hebe)	5

User 1 - Most Listened Genre

User Id	Genre Na..	
User 1	Pop	35
	UNK	27
	Metal/Rock	12

User/Audience

User dashboard
-   **Guide the listener to the top trends**

-   **Show what the listener has been listening to lately**

Most Popular Songs - Overall

周杰倫 (Jay Chou) 告白氣球	周杰倫 (Jay Chou) 不該	Eric 周興哲 你, 好不好？ (How Have You Been?)	謝和弦 (R-chord) 謝謝妳愛我 (Thanks For Your Love)	陳奕迅 (Eason Chan)
周湯豪 (NICKTHEREAL) 帥到分手		林俊傑 (JJ Lin) 不為誰而作的歌 (Twilight)		Alan Walker Faded
田馥甄 (Hebe) 小幸運 (A little happiness)		五月天 (Mayday) 後來的我們 (Here| After| Us)		

User Listened

10,315   13,616

User Id

User 2

Listens

15   147
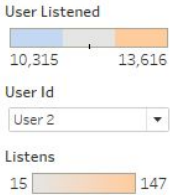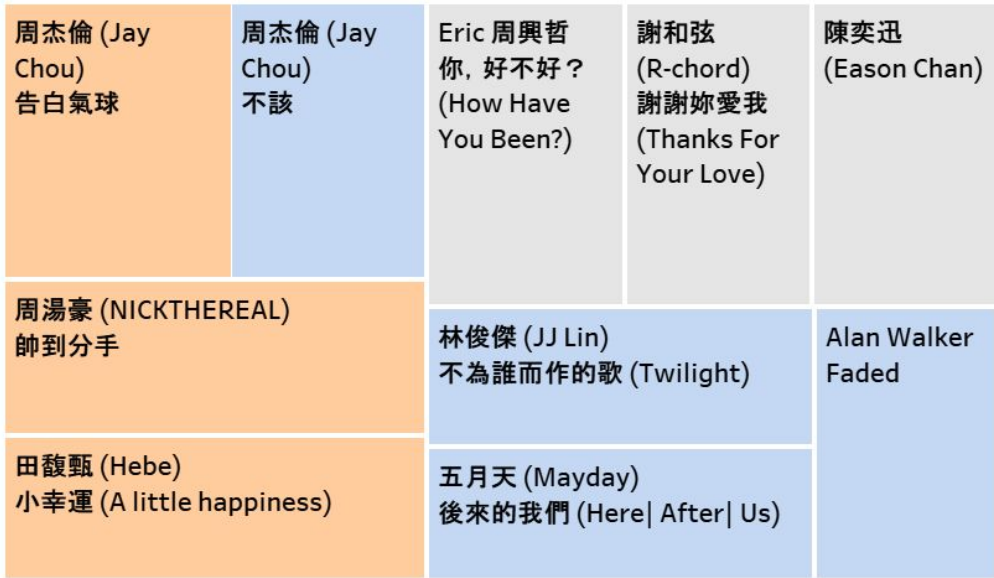
Most Popular Artists - Overall

Artist Name	
Various Artists	1
五月天 (Mayday)	2
周杰倫 (Jay Chou)	3
林俊傑 (JJ Lin)	4
田馥甄 (Hebe)	5

User 2 - Most Listened Genre

User Id	Genre Na..	
User 2	Pop	147
	Mandopop	38
	UNK	15
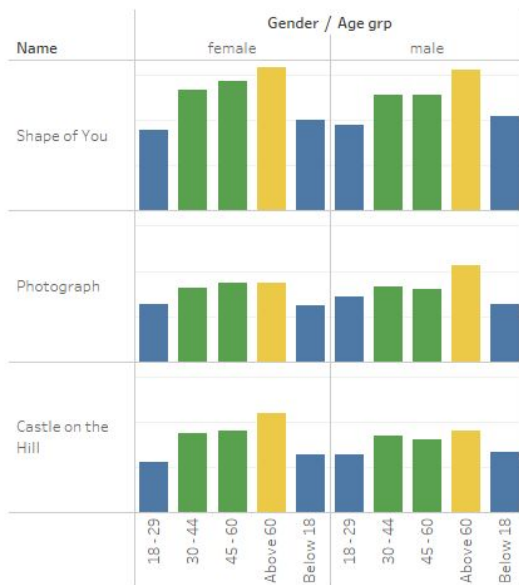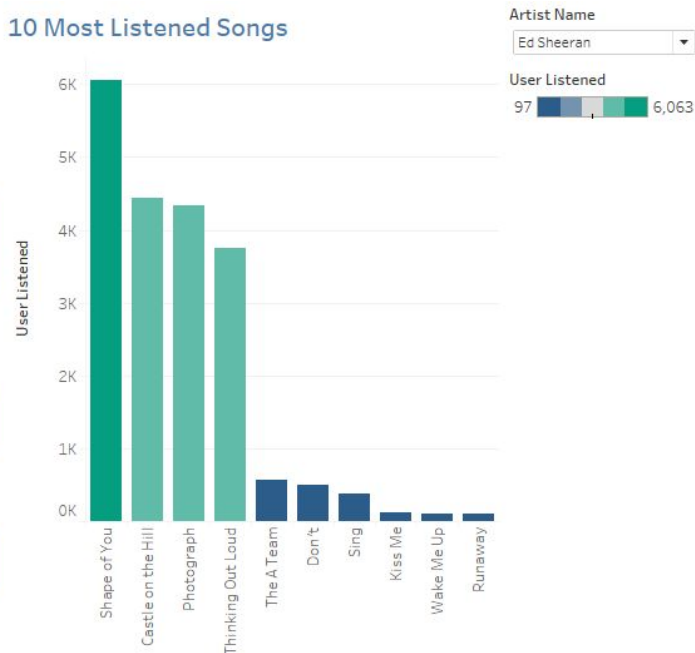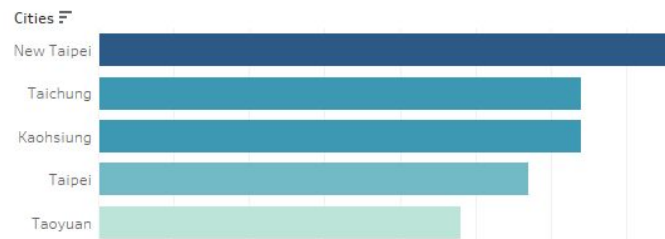
ML Song Likeness prediction model

We used the above **KKBOX** dataset to predict whether a user will like a song or not.

This was measured by whether the user listened to the said song again or not within a month from the song user relationship.

This was labeled as **target**.

We used **simple classification models** to predict this value using *7 different features*

The Song info and User info was combined with the song user relation dataset and used the shown features to predict the target value.

Features used
0:'song_length' (Song)
1:'genre_ids'(Song)
2:'language' (Song)
3: 'source_type' (Platform)
4: 'bd' (User Age)
5: 'city' (User)
6: 'gender' (User)

ML Song Likeness prediction model

The Target value ff a listener liked a song its 1, if not its 0.

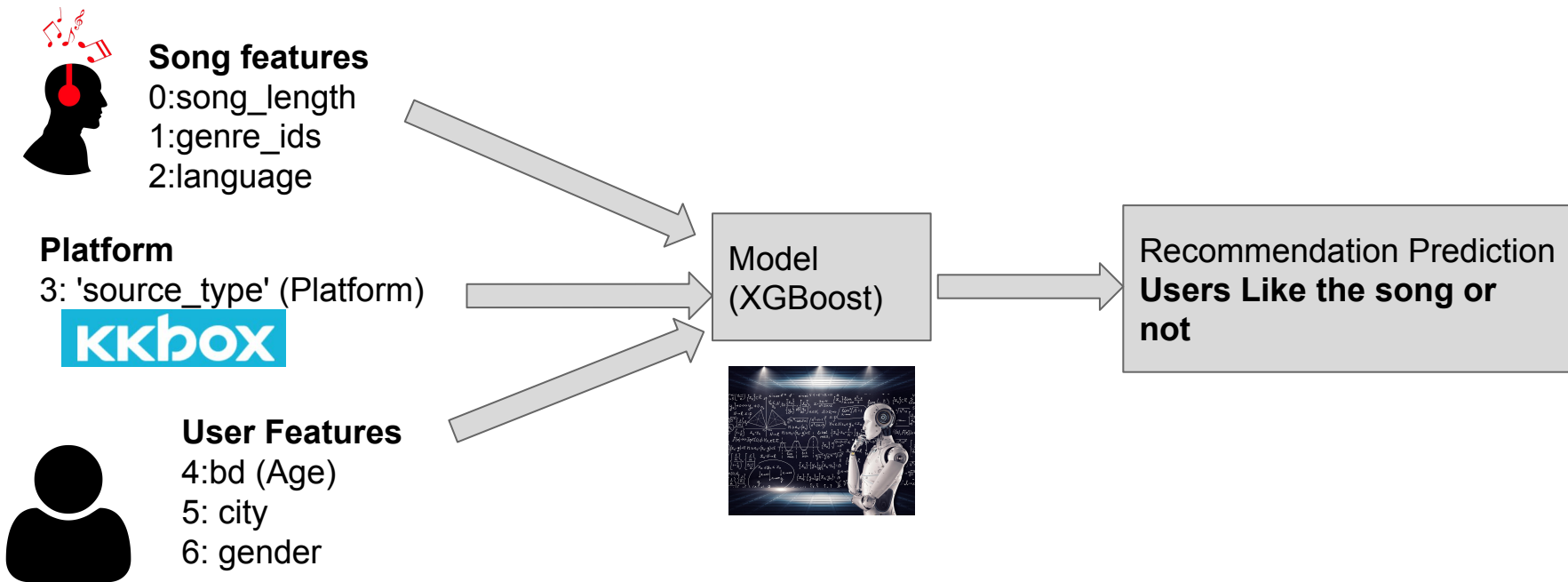The dataset is balanced with

0 : 0.49

1 : 0.51

The Train/Test split is 70:30

XGBoost got an accuracy of 64%. It could be improved if we had more and better features

Model	Test Accuracy (percent)
Logistic Regression	60
GaussianNB	57
XGBoost	*64*
Decision tree	59

ML Song Likeness prediction model

Song features
0:song_length
1:genre_ids
2:language

Platform
3: 'source_type' (Platform)

Model
(XGBoost)

Recommendation Prediction
Users Like the song or not

User Features
4:bd (Age)
5: city
6: gender

ML recommendation system

Song features
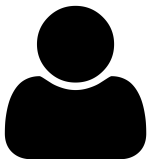0:song_length
1:genre_ids
2:language

Platform
3: 'source_type' (Platform)

KKBOX

User Features
4:bd (Age)
5: city
6: gender

So the system will take in 7 inputs that includes features from both user and the song and then predict whether the user will like the song or not.

We will use the XGBoost model that gives us 64% accuracy. We could obtain more accuracy if we had more features in our dataset.

ML Song Likeness prediction model

These are the available options for our model input of genre, language,source type and city

```
                         Program Instrucitons
_____
Fill the required information to predict the song will be popular for specific target user.

Available genre:
Mandopop, Pop, Pop Rock, Electronic, Progressive house, Jazz, K-Pop, Rap, Metal/Rock, Misc

Available language:
Taiwanese, Korean, Chinese, Japanese, Cantonese, Arabic, Thai, French, Unknown

Available source_type:
local-library , online-playlist, local-playlist, radio, album, top-hits-for-artist, song, listen-with,
song-based-playlist, topic-article-playlist, artist, my-daily-playlist

Available city:
New Taipei, Kaohsiung, Taichung, Taipei, Taoyuan, Tainan, Hsinchu, Keelung, Chiayi, Changhua, Pingtung, Zhubei,
Yuanlin, Douliu, Taitung, Hualien, Toufen, Nantou, Yilan, Miaoli, Magong

_____
```

ML Song Likeness prediction model

So the system will take in 7 inputs that includes features from both user and the song and then predict whether the user will like the song or not.

We will use the XGBoost model that gives us 64% accuracy. We could obtain more accuracy if we had more features in our dataset.

User Input:

```
Enter song_length(minutes and seconds) :
Enter song minutes : 2
Enter song seconds : 30
Enter genre:Pop
Enter language :Korean
Enter source_type :album
Enter Age :25
Enter city :Taipei
Enter gender :female
```

Predicted Result:

```
The user won't like this song!
```

ML Song Likeness prediction model

PREDICT LIKENESS

By being able to predict whether a user will like a song or not we can recommend songs to the user that maximises their listening probability hence increasing the platforms usage

Conclusion

Music streaming platforms aim for maximum customer engagement by increasing the duration of listening. This is what our BIS is aiming for.

This could be increased by analysing the data and providing various insights through visualisations and predictions which eventually help

- Recommending the right songs to the user,
- Showing which songs are performing well in the platform to the artists so they can continue creating those type of songs
- Showing overall high performance songs in the platform in order to promote these artists and songs to more listeners.

THANK YOU